Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 185 (2021) 282–291

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference Theme: Big Data, IoT, and AI for a Smarter Future
Malvern, Pennsylvania, June 16-18, 2021

Using Machine Learning and Wearable Inertial Sensor Data for the Classification of Fractal Gait Patterns in Women and Men During Load Carriage

Nizam U. Ahamed[a*], Kellen T. Krajewski[a], Camille C. Johnson[a], Adam J. Sterczala[a], Julie P. Greeves[b,c,d], Sophie L. Wardle[b,d], Thomas J. O'Leary[b,d], Qi Mi[a], Shawn D. Flanagan[a], Bradley C. Nindl[a], Chris Connaboy[a]

[a] *Neuromuscular Research Laboratory/Warrior Human Performance Research Center, Department of Sports Medicine and Nutrition, University of Pittsburgh, Pittsburgh, PA, USA*
[b] *Army Health and Performance Research, Army Headquarters, Ministry of Defence, UK*
[c] *Norwich Medical School, University of East Anglia, Norwich, UK.*
[d] *Division of Surgery and Interventional Science, University College London, London, UK*

Abstract

Ambulating while carrying a mission specific load is one of the most frequently executed occupational tasks for the military, especially for individuals in combat roles. Prolonged ambulation is a naturally dynamic and complex process, characterized by highly multi-dimensional interactions within the gait mechanics of the lower extremity. Recent wearable sensors studies, like inertial measurement unit (IMU)-related gait studies have demonstrated that machine learning (MLN) algorithms and fractal analysis can successfully discriminate between classes, such as movement patterns, injury, age and sex. This study attempts to classify fractal gait patterns of women and men using IMU-based signal data obtained from accelerometer, gyroscope and magnetometer during a 2 km loaded (20 kg) march. Random Forest (RF) MLN algorithm was used to generate a model that can measure the accuracy and identify the importance of IMU-based signal-related fractal variables. A total of 18 variables were calculated using 2 fractal methods, detrended fluctuation analysis (DFA) and wavelet transform-based power spectral density (PSD), from 3 IMU-based signals in their 3 axes (medial-lateral, vertical and anterior-posterior). A total of 33 healthy adults (17 men [26.7±5.9 years] and 16 women [25.2±4.5 years]) volunteered for this study. A 9-axis IMU sensor was attached to each participant at each of the following locations: feet, shanks, thighs and lumbar spine. An independent training-testing approach, called one-vs-one (*i.e.,* variables from one IMU-based signal were trained and tested using another IMU-based signal) was applied to determine the classification accuracy (*i.e.,* similarities between IMUs) and variable importance (score ranges: 0.0-1.0) measures. These values were then used to select the variables that best independently describe the rank in classification margin. The results from each IMU sensor placement based on the fractal values showed 'moderate' accuracy (50-75%), with the exception of two cases: the left shank yielded 'good' accuracy (80.1%) compared with the right shank, and the right thigh generated 'poor' accuracy (48.9%) compared with the left foot. No IMU location showed excellent accuracy (>90%). The results indicate that each IMU placement location has their own fractal patterns that are not similar to another IMU location in terms of sex classification. The analysis of the variable importance in the classification margin showed that most of the PSD resulted variables were classified as 'most important' compared with the DFA resulted variables. IMU sensors, and the associated analyses, could be used during military load carriage to evaluate changes in gait resulting from injury, fatigue or overtraining.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference, June 2021.
10.1016/j.procs.2021.05.030

Keywords: Gait; fractal; men and women; machine learning; load carriage; military

*Corresponding author: nizamahamed@pitt.edu

1. Introduction

Load carriage is a frequently performed task in the daily training and operations of a Warfighter, characterized by multi-dimensional interactions within their gait system [1, 2]. Such movement patterns during loaded marching constitute a continuous, cyclical activity, and although this activity is highly repeatable, loaded marching shows a subtle degree of variability in the stride-to-stride regulation of gait dynamics [2]. The magnitude and frequency of the gait variability is a potential predictor of fall risk and abnormal gait during loaded marching [3]. Musculoskeletal injuries (MSI) are prevalent among active-duty service members in military population and can result in gait modifications [4]. Ruck marching with load carriage has been implicated in the risk for acute and cumulative stress-induced MSI, and this risk may be exacerbated by prolonged duration of a loaded marching task [5]. Gait cycle stability and/or variability during loaded marching has been observed to be neither regular nor random, but shows a fractal temporal composition [6]. Whilst variability was once considered errant noise in the locomotor system, the subtle differences in dynamic structure of gait variability has demonstrated utility for the quantification of rehabilitation progression, monitoring normal/anomalous gait patterns, and/or monitoring changes in performance during training and operations [7, 8]. Therefore, it is important to assess the fractal dimension (i.e., more regular/self-similar) of the gait variability, which may be indicative of a more flexible/adaptive motor system, and relate that information to task experience (experts *vs.* trainees) and in healthy, compared with injured/fatigued warfighters during load carriage tasks [9]. In addition, recent sex-based gait studies have revealed that sex might have a substantial influence on gait mechanics during movement with standardized military-relevant symmetric loads [10]. Differences between women and men warfighters have been explored in the psychological, biological, and behavioral domains [11, 12], but the nature, patterns, and extent of the sex-based fractal dynamics of a warfighter's gait during load carriage remains mostly unexplored [6]. Thus, the measurement and classification of the variability structure (i.e., fractal dynamics) in warfighters' gait signal magnitude and frequency content during load carriage constitute an important area of research.

Over the last decade, many researchers have taken an interest in the application of multiple methods for the measurement of the fractal dynamics of gait variability. Two commonly used non-linear measures, detrended fluctuation analysis (DFA) and wavelet transform-based power spectral density (PSD), have been commonly applied to quantify the fractal patterns (regularity/self-similarity) of continuous and discrete gait time-series data [13]. DFA, which measures the strength of memory over time, and PSD, which addresses the frequency region of a time series, are two important fractal measuring techniques for gait time series that are used to scale the long-term autocorrelations of non-stationary signals [2]. These methods can quantify the variations in a time series by using its self-similar property and the fractal scaling index value, i.e., alpha (α) for DFA and beta (β) for PSD. The α and β values present an assessment of the statistical persistence or anti-persistence in gait time series data. An optimal value of α and β in gait is between 0.5 and 1.0 (where 1.0 indicates high persistence) [2]. This specifies the existence of statistical persistence between the stride intervals (i.e., pink noise or highly fractal), which indicates that the inter-stride-intervals between successive gait strides are constant (*i.e.,* non-random) at a long-range scale and that little deviations exist across multiple successive strides. In contrast, α and β values less than 0.5 indicate the existence of statistical anti-persistence between each step interval [8, 14, 15]. Recently, these two techniques have been used independently to evaluate and detect altered gait dynamics [13], but have not been applied sequentially to analyze the fractal patterns of gait during load carriage. Importantly, most gait dynamic studies are performed in a laboratory setting (e.g., treadmill and force plate) which can artificially influence gait variability, but there is a paucity of research investigating load carriage marching outside of a laboratory setting [7].

Wearable sensors allow the easy and uninterrupted measurements of body functions, such as physiological and biomechanical functions of the gait, in multiple environments [16]. Advances in technology, particularly those corresponding to small, inexpensive, and portable wearable sensors, such as inertial measurement unit (IMU), provide an alternative to laboratory-based gait research tools. These wearable sensors allow the potential for researchers to perform objective gait assessments of military populations, to better understand how individuals cope with the demands of the load carriage task and the diverse nature of the complex movement response associated with load carriage (*i.e.,* task specific perturbation) [1, 17]. IMUs incorporate built-in sensors, such as an accelerometer (measuring acceleration), a gyroscope (measuring angular velocity), and a magnetometer (measuring magnetic orientation), along three axes, namely, the anterior-posterior (AP), vertical (VT) and medial-lateral (ML) [18]. A previous gait-related study revealed that the fractal patterns during non-military activities differ based on the IMU placement locations and their associated axes [19]. Therefore, these IMU generated signals along the three orthogonal axes may be useful for the quantification of fractal gait patterns and the comparison of gait patterns between men and women. To date, these signals have not been adequately studied during load carriage tasks, due to the large complex nature of the datasets and a lack of accessibility to a portable sensing system.

With the increased accessibility to higher processing power computers, data-driven approaches, such as machine learning (MLN) algorithms, have shown promise when applied to IMU generated gait data for handling complex patterns from data in human movement. However, only a handful of studies have explored using both IMU and MLN algorithms together in military-related operational tasks and injury prediction [20, 21], and none have explored load carriage performance. Among the variety of MLN algorithms, random forest (RF) is one of the most widely used non-linear ensemble-based supervised MLN algorithms, and is utilized to perform classification tasks with good and reliable accuracy from large-scale datasets [22]. Additionally, RF can solve the complex problem of variable selection through the enhanced identification of the most important variable(s) for improving the prediction accuracy of recognizing normal and/or anomalous gait events [22]. A number of IMU-based gait related studies aimed to perform classification using the RF classifier, such as predicting the age and sex of the individuals, the typical and anomalous gait patterns and other classification-related tasks [23, 24]. A few studies investigated non-military related load carriage activities [25, 26], but to-date, fractal-related variables have not been used for the classification of data obtained during military-related activities.

An MLN-based approach classifying and comparing fractal gait patterns between sexes from IMU-generated signals during load carriage outside of a laboratory setting has not previously been established. Female soldiers have a higher risk of developing MSI compared with their male counterparts, and anomalous marching gait patterns during load carriage are considered an important factor in MSI risk [27, 28]. However, no study has explored sex differences in gait complexity, exploiting IMU placement at lower extremity in relation to observed gait complexity during load carriage. To address this important research gap, two types of fractal measuring techniques (DFA and PSD) and the MLN algorithm RF were applied to answer the following research questions: 1) "Does the fractal patterns (based on the accuracy [%] results) of men and women with the same IMU placement location vary from those obtained with another IMU placement location?", and 2) "What variable(s) is/are most important for the classification margin?". To accomplish this goal, we developed and described an MNL-based automated system that can classify and separate subjects by sex using a set of gait fractal variables. We hypothesized that each IMU placement location is associated with specific fractal characteristics, and as a result, less similarity would be observed among IMUs in terms of the classification of women and men, which could be reflected by a decreased classification accuracy. Second, PSD (frequency characteristic)-related variables will be more important than DFA-related variables in the classification margin for separating men and women.

2. Methods

2.1. Participants

A total of 33 recruit-aged healthy adults (16 women and 17 men: [mean±SD], age = 25.2 ± 4.5 and 26.7 ± 5.9 years, %body fat = 32.1 ± 6.3 and 20.6 ± 5.9, body mass = 66.9 ± 11.8 and 75.6±11.1 kg, height = 165.3 ± 5.6 and 178.8 ± 7.4 cm, respectively) participated and gave written informed consent before commencing the experiment. The study was approved by the University of Pittsburgh, USA institutional review board and UK Ministry of Defence research ethics committees.

2.2. Experimental protocol

Each participant completed the physical employment standard (PES) tests for British Army Ground Close Combat (GCC) roles. The PES tests are a series of physical military occupational tasks designed to represent the mechanical and physiological requirements of GCC employment. Gait data from 7 IMU locations on each participant were collected during their 2 km best effort loaded march which was performed outside of the laboratory setting. The 20 kg load was composed of a 10 kg plate carrier, an 8 kg load, a 1.5 kg simulation rifle and a 0.5 kg helmet. The participant load carriage configuration is illustrated in Figure 1. The rucksack was packed with a 7 kg crumb-rubber pack (inferior portion of the pack) and light pillows to ensure that the pack was filled and that its weight was evenly distributed to reduce asymmetrical load perturbations. The entire testing procedures have been described in our earlier published article [6].

2.3. Instrumentation

Data were captured using a wearable 9-axis IMU sensor (IMeasureU Blue Thunder Ltd.) and the sampling rate was 500 Hz. The IMU sensor contains a tri-axial accelerometer (± 16g), a tri-axial gyroscope (2000°/s), and a tri-axial magnetometer (± 1200 μT) consisting of acceleration, angular velocity (rotational motion), and magnetic field vector or magnetic flux (magnetism) data, respectively, in the three orthogonal planes, namely, the vertical (*y* axis), the medial-lateral (*x* axis), and the anterior-posterior (*z* axis) planes.

A total of 7 individual IMUs were employed to collect accelerometer, gyroscope, and magnetometer characteristics for the lower body. The IMUs were attached at the following 7 locations: right and left thighs, right and left shanks, right and left feet, and the lumbar spine (Figure 1) [6]. If the clothing was tight fitting (*i.e.*, compression gear), the IMUs were placed over the clothing; if the participant wore loose-fitting clothing (*i.e.*, basketball shorts), the IMUs were placed directly on the skin underneath the clothing to reduce movement artifact from the fabric. The lumbar spine IMU was always placed directly on the skin.

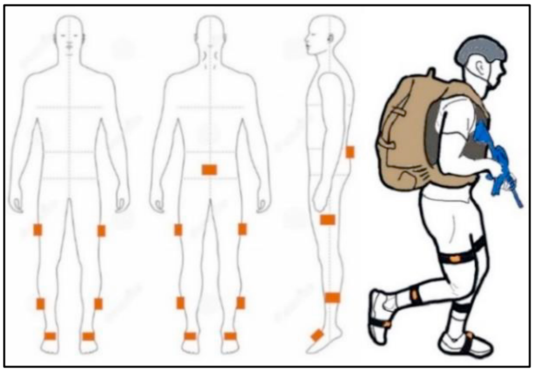

Fig. 1. IMU placement during load carriage

2.4. Signal Processing

The continuous gait time series data from the accelerometer, gyroscope, and magnetometer were processed for analysis. The recorded signal was low pass filtered (4^{th}-order Butterworth) with a cut-off frequency (*fc*) of 18 Hz. Each participant completed the 2 km distance loaded march; the first- and last- 30 seconds of each participant's time series data were removed. The resulted cleaned and filtered data were then processed for the feature extraction purpose using two types of fractal scaling methods and for the classification purpose using the RF MLN algorithm. DFA and PSD were used to assess the fractal values of the gait time-series from 7 IMU locations of men and women to quantify gait dynamics by the outcome measure alpha (α) and beta (β) values, respectively.

DFA is a widely used method for the analysis of fractal patterns [29]. Wavelet-based PSD assessment method provides important information on how power (i.e., variance) is distributed as a function of frequency. Additionally, this is an appropriate technique for the analysis of human gait-related signals. The PSD β-values can evaluate the presence of long-range power law correlation in time series data for calculating the fractal dimension [30]. Calculation of the PSD is a frequently-used analytic technique for defining periodicities in time series [31]. The α- and β-values were calculated from the ~27-min duration of the continuous time-series data (the IMU data were binned into smaller data segments using 60-s sliding windows, *i.e.,* each 30,000 consecutive continuous data points segment) of each participant. A total of 18 variables were yielded using DFA (α) and PSD (β) methods: (*i.e.,* 2 fractal methods × 3 types of signal × 3 axes = 18 variables) to create an IMU-based dataset for the training and testing purpose of the MLN algorithm. Table 1 presents the name of variables acquired from DFA and PSD methods that were utilized to differentiate fractal patterns of gait in men and women. The entire signal processing tasks were performed using custom-made software in MATLAB (2019b; The MathWorks Inc. USA).

Table 1. Variables generated using two fractal methods (DFA and PSD) for the analysis of three signal types in three axes.

Using detrended fluctuation analysis (DFA)			**Using power spectral density (PSD)**		
Var#	**Acronyms**	**Meaning**	**Var#**	**Acronyms**	**Meaning**
D1.	D_acc_AP	DFA α of acceleration at AP axis	P1.	P_acc_AP	PSD β of acceleration at AP axis
D2.	D_acc_VT	DFA α of acceleration at VT axis	P2.	P_acc_VT	PSD β of acceleration at VT axis
D3.	D_acc_ML	DFA α of acceleration at ML axis	P3.	P_acc_ML	PSD β of acceleration at ML axis
D4.	D_gyr_AP	DFA α of AV at AP axis	P4.	P_gyr_AP	PSD β of AV at AP axis
D5.	D_gyr _VT	DFA α of AV at VT axis	P5.	P_gyr _VT	PSD β of AV at VT axis
D6.	D_gyr _ML	DFA α of AV at ML axis	P6.	P_gyr _ML	PSD β of AV at ML axis
D7.	D_mag_AP	DFA α of MFV at AP axis	P7.	P_mag_AP	PSD β of MFV at AP axis
D8.	D_mag _VT	DFA α of MFV at VT axis	P8.	P_mag _VT	PSD β of MFV at VT axis
D9.	D_mag _ML	DFA α of MFV at ML axis	P9.	P_mag _ML	PSD β of MFV at ML axis

acc: accelerometer, gyr: gyroscope, mag: magnetometer; AV: angular velocity, MFV: magnetic field vector, AP: anterior-posterior (z axis), VT: vertical (y axis), ML: medial-lateral (x axis)

2.5. Machine learning:

The RF-based predictive model measures classification accuracy and variable importance from the IMU-generated dataset. The classification task was done by assessing the performance indicators or scores to calculate the classification by calculating total accuracy in percentage (%) and specified by the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) (equation 1) [32].

$$\text{Accuracy (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

To build and run (train and test) the RF-based model, the Python programming language (www.python.org, Python Software Foundation, USA) was used using its built-in Anaconda distribution. Some of the notable packages used to write the code in Python programming include scikit-learn, matplotlib, scipy and numpy [22]. Some important selected parameters were *max_depth* (maximum depth of the tree), *max_features* (number of features selected to split each node) and *n_estimators* (total trees in the forest [*e.g.,* 50, 100]). These parameters were fine-tuned to improve the total classification accuracy. The accuracies were interpreted as, excellent: >90%, good: 75-90%, moderate: 50-75%, and poor: <50% [18]. The test criteria between two IMU datasets were established using the developed RF-based model to classify the fractal patterns of gait from the training data correctly. We used a one-*vs*-one approach to ensure the independent training and testing in the classification, *i.e.,* data from one IMU were trained and tested with another IMU (*e.g.,* right *vs.* left thigh signal data) [18]. Furthermore, in order to build the decision trees, the RF algorithm automatically used a Gini-index to assess the impurity of a node from the classification and regression tree learning system [33]. These RF-based trees calculate the importance of the fractal-based variables to categorize sexes using its in-built feature importance measures. In general, variable importance presents a score (ranges from 0.0 to 1.0) that indicates how useful each variable (*i.e.,* influence of each variable in the classification margin) was in the building of the RF trees within the model and ultimately the sum of all the variables is equal to 1.0.

3. Results

Table 2 shows results for classification accuracy generated with the one-*vs*-one training-testing approach. As shown in the table, all the individual results show 'moderate' accuracies (*i.e.*, values in the range of 50 to 75%) with the exception of two cases (*i.e.,* the left shank generated 'better' accuracy [80.1%] compared with the right shank, and the right thigh generated 'poor' accuracy [48.9%] compared with the left foot). 'Moderate' accuracy was also found when the overall mean values were calculated from the test dataset, and in this dataset, the shanks exhibited the highest accuracies (65.6% for the right shank and 65.5% for the left shank) than the other IMU placements. Notably, we did not consider the results with 100% accuracies generated with the dependent dataset, *i.e.*, the training-testing datasets were obtained from the same IMU.

Table 2: Experimental results (classification accuracy and most important variable) using one-*vs*-one training-testing approach.

Training ↓ / Testing →	**Left foot**	**Left shank**	**Left thigh**	**Right foot**	**Right shank**	**Right thigh**	**Lumber spine**
Left Foot	**100.0 (P7)**	60.5 (P7)	60.0 (P7)	65.8 (P7)	70.1 (P7)	48.9 (P7)	72.8 (P7)
Left Shank	64.4 (P1)	**100.0 (P1)**	58.6 (P1)	66.5 (P1)	74.1 (P1)	56.3 (P1)	57.2 (P1)
Left Thigh	64.0 (P4)	64.2 (P6)	**100.0 (P4)**	67.0 (P4)	63.8 (P4)	72.4 (P7)	58.2 (P8)
Right Foot	69.9 (P1)	72.0 (P5)	59.5 (P5)	**100.0 (P5)**	64.8 (P5)	58.6 (P1)	62.8 (P9)
Right Shank	61.1 (P1)	80.1 (P1)	63.3 (P1)	68.8 (P1)	**100.0 (P1)**	58.3 (P1)	61.4 (P1)
Right Thigh	53.1 (P1)	57.9 (P7)	72.5 (P4)	61.9 (P5)	57.8 (P6)	**100.0 (P4)**	56.8 (P5)
Lumbar spine	58.1 (P2)	57.9 (P2)	54.9 (P7)	62.4 (P4)	63.2 (P2)	55.9 (P2)	**100.0 (P2)**
***Average**	**61.7**	**65.5**	**61.5**	**65.4**	**65.6**	**58.4**	**61.6**

* To calculate the average accuracies, the dependent training-testing criteria was excluded, *i.e.,* training-testing was conducted using the same dataset and resulted in **100%** accuracies (**black bolded font**). So, only the independent accuracies, *i.e.*, different training and testing datasets, are included.

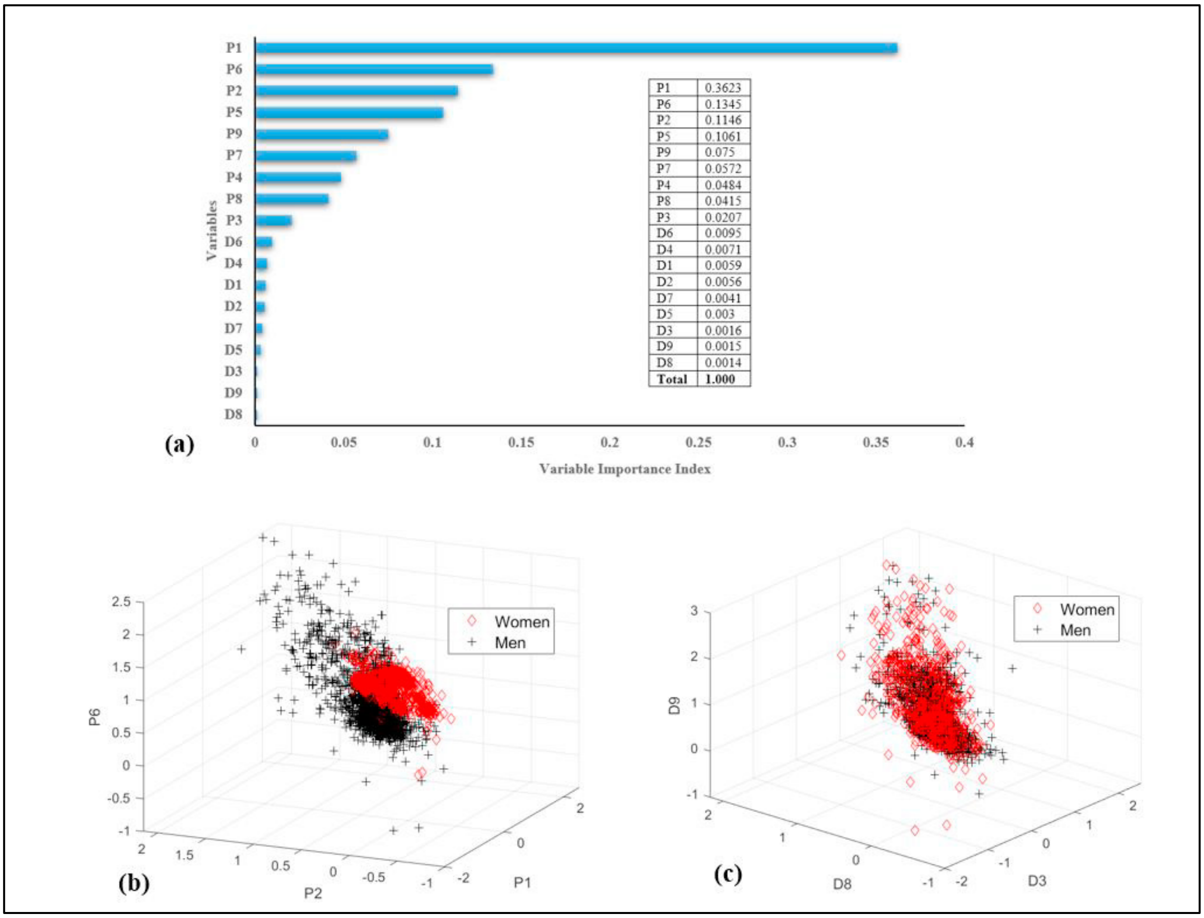

Fig. 2. Most important variables in the gender classification margin obtained by training with data from the right shank and testing with the data from the left shank that generated the highest classification accuracy (80%). (a) The different features are ranked by their relative importance, and the feature importance is normalized such that the various values add up to 1.0. An example of 3D plots for sex differences according to (b) the three most important variables (P1, P6 and P2) and (c) the three least important variables (D8, D9 and D3).

The variable importance in the classification margin are displayed in the Table 2, and the variable names are displayed in parentheses immediately after the classification accuracy. The results show that no DFA variables were found as 'most important' in the classification task. In contrast, all PSD-based variables with the exception of the variable P3 were classified as 'most important' and contributed to the separation of men and women. P1 (*i.e.*, PSD values for the acceleration in the AP axis) were most frequently identified (17 times) as the most important variable, followed by P7, P4, P5, P2, P6, P8 and P9, which were identified 10, 7, 6, 5, 2, 1 and 1 times, respectively. Figure 2(a) illustrates an example of how the important variables contribute to the RF-based classification model obtained using the data from the right shank as the trained dataset and the data from the left shank as the test dataset. The two graphical three-dimensional (3-D) plots shown in Figures 2(b) and 2(c) present the three most-important variables (P1, P6 and P2) and the three least-important variables (D8, D9 and D3), respectively. These 3-D plots clearly present how the variables contribute to the separation of men and women in the classification margin.

4. Discussion

This study aimed to employ wearable IMU-based signal dataset and the MLN approach for the classification of gait fractality (using PSD and DFA) in women and men. Secondarily, this study evaluated whether the fractal gait patterns obtained during load carriage from different IMUs placed in the lower extremity differed between women and men. Two main novel findings were obtained: 1) each IMU location is associated with a specific fractal pattern, as demonstrated by the 'moderate' classification accuracies (*i.e*., moderate similarities) achieved with the IMUs, and 2) PSD (characteristic of the frequency region)-related variables are better predictors than DFA-related variables in the classification margin. We also demonstrated that wearable sensors can successfully capture gait data and that RF-based MLN can discriminate gait complexity from these data, an area which may have been ignored in previous gait studies performed using traditional data processing applications. These findings demonstrate the existence of further opportunities for advancing science, clinical care, and warfighter's health through the development of tools for detecting an anomalous gait structure that might result from injury, fatigue or overtraining outside of the laboratory setting.

Our MNL and inter-IMU based fractal patterns (*i.e.,* classification accuracy) identification are in agreement with previous IMU-based findings regarding the use of gait pattern differences (*e.g.,* step length, step width, velocity and acceleration) across sensor locations for classifying various activities/conditions (*e.g.,* women *vs.* men, healthy *vs.* patient, young *vs.* old, osteoarthritis *vs*. non-osteoarthritis and falling *vs.* not falling [34-36]. These studies showed that gait patterns (*e.g.,* complexity) may be influenced by IMU position and placement locations, target variables for classification and computational approaches. Although IMUs can be easily worn by individuals, the magnitude to which the placement/positioning of an IMU as part of the protocol can influence the accuracy and utility of the measurements collected during load carriage is unknown. These findings highlight the importance of assessing gait complexity based on each IMU placement. Because computational intelligence (*e.g.,* MLN)-assisted wearable - based automated systems have become increasingly ubiquitous during everyday activities of daily living [18], the possibilities of enhancing human-computer interaction could be boundless [37].

The RF-based MLN algorithm has achieved growing interest in gait related, cross-sectional studies due to its capability to assess the importance of a variable in yielding good classification performance and fast-computational processing speed [18]. Our RF-based model shows that DFA values contribute less (*i.e*., are less important) to the sex classification than the PSD values. A possible explanation for this finding is the variation in the peak frequency in the spectral profile during movement which changes more frequently than other spatial and temporal gait parameters in both women and men [38]. PSD β-values of acceleration signal at AP axis is less stable than the other two axes (i.e., ML and VT), and the PSD-related variables in the frequency spectrum of acceleration in the AP axis is a more appropriate tool for measuring gait timing variables (i.e., stride time), rather than ML and VT axes [6]. Secondly, several DFA results from IMU signals failed to find significant differences between sexes, mostly due to the fact that no difference have been detected in DFA-related variables between women and men [39]. Thus, our IMU-based results suggest that PSD variables (β-values) are better contributors to sex classification than DFA-related variables (α-values) during load carriage. This finding may be apparent as a result of the spectral frequency component of the gait signals (i.e., small/large spectral peaks) are varied between women and men.

A limitation of this study is that continuous gait time series dataset rather than stride-to-stride interval time series were used to calculate the fractal values. However, many studies have used continuous time series data, without counting the stride interval and considered only the epoch time length to calculate the fractal dimension [40, 41]. Another limitation is that the scaling values (α and β values) were not reported directly in the paper, but we used all of these values in the MLN algorithm to calculate the accuracy (similarities). In future research, more associated metrics (*e.g.,* heart rate, sleep index and global positioning system data) from the wearable sensors in addition to gait variables should be used to monitor the gait complexity, injury risk and performance. Additionally, more advanced model-based artificial intelligence algorithms (e.g., fuzzy expert system) could be used to assess the large and heterogeneous datasets from wearable systems to recognize complex gait patterns during military load carriage task [42].

In conclusion, our wearable IMU-specific MLN results (*i.e.,* classification accuracy and variable importance measure) provide new insight into the fractal differences apparent in women and men performing load carriage tasks. Additionally, the findings based on wearable inertial sensor data for the classification of fractal gait patterns between women and men further emphasize the significance of considering sex as an important predictor variable in studies of military gait during load carriage. These MLN- and IMU-based cross-sectional results are the first step in gait classification which could lead to more sensitive measures to classify individuals based on MSI risk.

Acknowledgements

The study was supported by the US Dept. of Defense (W81XWH18SBAA1).

References

[1] Shi, H., Zhao, H., Liu, Y., Gao, W., and Dou, S.-C. (2019) "Systematic Analysis of a Military Wearable Device Based on a Multi-Level Fusion Framework: Research Directions." *Sensors*, **19, (12)**, 2651.

[2] Krajewski, K.T., Dever, D.E., Johnson, C.C., Mi, Q., Simpson, R.J., Graham, S.M., Moir, G.L., Ahamed, N.U., Flanagan, S.D., and Anderst, W.J. (2020) "Load Magnitude and Locomotion Pattern Alter Locomotor System Function in Healthy Young Adult Women." *Frontiers in Bioengineering and Biotechnology*, **8**, 1066.

[3] Myung, R., and Smith, J.L. (1997) "The Effect of Load Carrying and Floor Contaminants on Slip and Fall Parameters." *Ergonomics*, **40, (2)**, 235-246.

[4] Wardle, S.L., and Greeves, J.P. (2017) "Mitigating the Risk of Musculoskeletal Injury: A Systematic Review of the Most Effective Injury Prevention Strategies for Military Personnel." *Journal of Science and Medicine in Sport*, **20**, S3-S10.

[5] Poel, D.N. (2016) "The Effects of Military Style Ruck Marching on Lower Extremity Loading."

[6] Ahamed, N.U., Krajewski, K.T., Johnson, C.C., Sterczala, A.J., Greeves, J.P., Wardle, S.L., O'Leary, T.J., Mi, Q., Flanagan, S.D., and Nindl, B.C. (2020) "Using Wavelet-Based Fractal Analysis of Inertial Measurement Unit Signals to Examine Gait Data from Men and Women During a Load Carriage Task." *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*.

[7] Mahzoun Alzakerin, H., Halkiadakis, Y., and Morgan, K.D. (2020) "Characterizing Gait Pattern Dynamics During Symmetric and Asymmetric Walking Using Autoregressive Modeling." *PloS one*, **15, (12)**, e0243221.

[8] Krajewski, K.T., Dever, D.E., Johnson, C.C., Rawcliffe, A.J., Ahamed, N.U., Flanagan, S.D., Mi, Q., Anderst, W.J., and Connaboy, C.. (2020) "Load Carriage Magnitude and Locomotion Strategy Alter Knee Total Joint Moment During Bipedal Ambulatory Tasks in Recruit-Aged Women." *Journal of Biomechanics*, 109772.

[9] Knapik, J.J., Canham-chervak, M., Hauret, K., Laurin, M.J., Hoedebecke, E., Craig, S., and Montain, S.J. (2002) Seasonal Variations in Injury Rates During Us Army Basic Combat Training." *Annals of occupational hygiene*, **46, (1)**, 15-23.

[10] Loverro, K.L., Hasselquist, L., and Lewis, C.L. (2019) "Females and Males Use Different Hip and Knee Mechanics in Response to Symmetric Military-Relevant Loads." *Journal of Biomechanics*, **95**, 109280.

[11] Varley-Campbell, J., Cooper, C., Wilkerson, D., Wardle, S., Greeves, J., and Lorenc, T. (2018) "Sex-Specific Changes in Physical Performance Following Military Training: A Systematic Review." *Sports Medicine*, **48, (11)**, 2623-2640.

[12] Lin, Y.-K., Lin, Y.-P., Lee, J.-T., Lin, C.-S., Wu, T.-J., Tsai, K.-Z., Su, F.-Y., Kwon, Y., Hoshide, S., and Lin, G.-M. (2020) "Sex-Specific Association of Hyperuricemia with Cardiometabolic Abnormalities in a Military Cohort: The Chief Study." *Medicine*, **99, (12)**, p. e19535.

[13] Warlop, T., Detrembleur, C., Stoquart, G., Lejeune, T., and Jeanjean, A. (2018) "Gait Complexity and Regularity Are Differently Modulated by Treadmill Walking in Parkinson's Disease and Healthy Population." *Frontiers in Physiology*, **9**, 68.

[14] Hausdorff, J., Zemany, L., Peng, C.-K., and Goldberger, A. (1999) "Maturation of Gait Dynamics: Stride-to-Stride Variability and Its Temporal Organization in Children." *Journal of Applied Physiology*, **86, (3)**, 1040-1047.

[15] Moumdjian, L., Maes, P.-J., Dalla Bella, S., Decker, L.M., Moens, B., Feys, P., and Leman, M. (2020) "Detrended Fluctuation Analysis of Gait Dynamics When Entraining to Music and Metronomes at Different Tempi in Persons with Multiple Sclerosis." *Scientific Reports*, **10, (1)**, 1-12.

[16] Li, X., Dunn, J., Salins, D., Zhou, G., Zhou, W., Schüssler-Fiorenza Rose, S.M., Perelman, D., Colbert, E., Runge, R., and Rego, S. (2017) "Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information." *PLoS Biology*, **15, (1)**, e2001402.

[17] Park, S.Y., Ju, H., and Park, C.G. (2016) "Stance Phase Detection of Multiple Actions for Military Drill Using Foot-Mounted Imu." *Sensors*, **14**, 16.

[18] Ahamed, N.U., Kobsar, D., Benson, L., Clermont, C., Kohrs, R., Osis, S.T., and Ferber, R. (2018) "Using Wearable Sensors to Classify Subject-Specific Running Biomechanical Gait Patterns Based on Changes in Environmental Weather Conditions." *PloS One*, **13, (9)**, e0203839.

[19] Shi, L., Duan, F., Yang, Y., and Sun, Z. (2019) "The Effect of Treadmill Walking on Gait and Upper Trunk through Linear and Nonlinear Analysis Methods." *Sensors*, **19, (9)**, 2204.

[20] Leightley, D., Williamson, V., Darby, J., and Fear, N.T. (2019) "Identifying Probable Post-Traumatic Stress Disorder: Applying Supervised Machine Learning to Data from a Uk Military Cohort." *Journal of Mental Health*, **28, (1)**, 34-41.

[21] Papadakis, N., Havenetidis, K., Papadopoulos, D., and Bissas, A. (2020) "Employing Body-Fixed Sensors and Machine Learning to Predict Physical Activity in Military Personnel." *BMJ Mil Health*.

[22] Ahamed, N.U., Kobsar, D., Benson, L.C., Clermont, C.A., Osis, S.T., and Ferber, R. (2019) "Subject-Specific and Group-Based Running Pattern Classification Using a Single Wearable Sensor." *Journal of Biomechanics*, **84**, 227-233.

[23] Han, Y.C., Wong, K.I., and Murray, I. (2019) "Gait Phase Detection for Normal and Abnormal Gaits Using Imu." *IEEE Sensors Journal*, **19, (9)**, 3439-3448.

[24] Garofalo, G., Argones Rúa, E., Preuveneers, D., and Joosen, W. (2019) "A Systematic Comparison of Age and Gender Prediction on Imu Sensor-Based Gait Traces." *Sensors*, **19, (13)**, 2945.

[25] Goršič, M., Dai, B., and Novak, D. (2020) "Load Position and Weight Classification During Carrying Gait Using Wearable Inertial and Electromyographic Sensors." *Sensors*, **20, (17)**, 4963.

[26] Lee, H., Yang, K., Kim, N., and Ahn, C.R. (2020) "Detecting Excessive Load-Carrying Tasks Using a Deep Learning Network with a Gramian Angular Field." *Automation in Construction*, 120, 103390.

[27] Epstein, Y., Fleischmann, C., Yanovich, R., and Heled, Y. (2015) "Physiological and Medical Aspects That Put Women Soldiers at Increased Risk for Overuse Injuries." *The Journal of Strength & Conditioning Research*, **29**, S107-S110.

[28] Orr, R.M., Pope, R., Johnston, V., and Coyle, J. (2014) "Soldier Occupational Load Carriage: A Narrative Review of Associated Injuries." *International Journal of Injury Control and Safety Promotion*, **21, (4)**, 388-396.
[29] Peng, C.K., Havlin, S., Stanley, H.E., and Goldberger, A.L. (1995) "Quantification of Scaling Exponents and Crossover Phenomena in Nonstationary Heartbeat Time Series." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **5, (1)**, 82-87.
[30] Sekine, M., Tamura, T., Akay, M., Fujimoto, T., Togawa, T., and Fukui, Y. (2002) "Discrimination of Walking Patterns Using Wavelet-Based Fractal Analysis." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **10, (3)**, 188-196.
[31] Moody, G.B. (1993) "Spectral Analysis of Heart Rate without Resampling." *Proceedings of Computers in Cardiology Conference*, (IEEE).
[32] O'Reilly, M.A., Whelan, D.F., Ward, T.E., Delahunt, E., and Caulfield, B. (2017) "Classification of Lunge Biomechanics with Multiple and Individual Inertial Measurement Units." *Sports Biomechanics*, **16, (3)**, 342-360.
[33] Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010) "Variable Selection Using Random Forests." *Pattern Recognition Letters*, **31, (14)**, 2225-2236.
[34] Sharifi Renani, M., Myers, C.A., Zandie, R., Mahoor, M.H., Davidson, B.S., and Clary, C.W. (2020) "Deep Learning in Gait Parameter Prediction for Oa and Tka Patients Wearing Imu Sensors." *Sensors*, **20, (19)**, 5553.
[35] Ahad, M.A.R., Ngo, T.T., Antar, A.D., Ahmed, M., Hossain, T., Muramatsu, D., Makihara, Y., Inoue, S., and Yagi, Y. (2020) "Wearable Sensor-Based Gait Analysis for Age and Gender Estimation." *Sensors*, **20, (8)**, 2424.
[36] Panebianco, G.P., Bisi, M.C., Stagni, R., and Fantozzi, S. (2018) "Analysis of the Performance of 17 Algorithms from a Systematic Review: Influence of Sensor Position, Analysed Variable and Computational Approach in Gait Timing Estimation from Imu Measurements." *Gait & Posture*, **66**, 76-82.
[37] Figueiredo, J., Santos, C.P., and Moreno, J.C. (2018) "Automatic Recognition of Gait Patterns in Human Motor Disorders Using Machine Learning: A Review." *Medical Engineering & Physics*, **53**, 1-12.
[38] Ferster, M.L., Mazilu, S., and Tröster, G. (2015) "Gait Parameters Change Prior to Freezing in Parkinson's Disease: A Data-Driven Study with Wearable Inertial Units." *Proceedings of the 10th EAI International Conference on Body Area Networks*.
[39] Pollind, M., and Soangra, R. (2020) "Development and Validation of Wearable Inertial Sensor System for Postural Sway Analysis." *Measurement*, **165**, 108101.
[40] Sosnoff, J.J., Goldman, M.D., and Motl, R.W. (2010) "Real-Life Walking Impairment in Multiple Sclerosis: Preliminary Comparison of Four Methods for Processing Accelerometry Data." *Multiple Sclerosis Journal*, **16, (7)**, 868-877.
[41] Buzzi, U.H., Stergiou, N., Kurz, M.J., Hageman, P.A., and Heidel, J. (2003) "Nonlinear Dynamics Indicates Aging Affects Variability During Gait." *Clinical Biomechanics*, **18, (5)**, 435-443.
[42] Ahamed, N.U., Benson, L., Clermont, C., Osis, S.T., and Ferber, R. (2017) "Fuzzy Inference System-Based Recognition of Slow, Medium and Fast Running Conditions Using a Triaxial Accelerometer." *Procedia Computer Science*, **114**, 401-407.